

Revelation of Anti-Termination Mechanism - Through Structural and Functional Analyses of HutP

Thirumananseri Kumarevel*

RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo 679-5148, Japan

Abstract: Regulating gene expression directly at the mRNA level represents a novel approach in the control of cellular processes in all organisms. In this respect, RNA-binding proteins, while in the presence of their cognate ligands, play a key role by targeting the mRNA to regulate its expression through attenuation or anti-termination mechanisms. Although many proteins are known to use these mechanisms in the regulation of gene expression, no structural insights have been revealed, to date, to explain how these proteins trigger the conformation for the recognition of RNA. This review describes the HutP mediated anti-termination mechanism by combining the *in vivo*, *in vitro* and X-ray analyses of the activated conformation of HutP, initiated by the coordination of L-histidine and Mg^{2+} ions, based on our previous and recently solved crystal structures (uncomplexed HutP, HutP- Mg^{2+} , HutP-L-histidine, HutP- Mg^{2+} -L-histidine, HutP- Mg^{2+} -L-histidine-RNA (21-mer and 55-mer)). In this anti-termination process, HutP initiates destabilization at the 5'-end of its mRNA by binding to the first UAG-rich region and then accesses the second UAG-rich region, located downstream of the stable G-C-rich segment of the terminator stem. By this mode of action, HutP appears to disrupt the G-C rich terminator stem loop, and allow the RNA polymerase to pass through the destabilized terminator, thus it prevents premature termination of transcription in the RNA segment preceding the regions encoding for the genes responsible for histidine degradation.

Keywords: Anti-termination, attenuation, *hutp*, RNA-binding proteins, metal ions, allosteric activation, transcription regulation.

INTRODUCTION

Bacteria exploit a variety of mechanisms to regulate transcription elongation in order to control gene expression in response to changes in their environment. A frequently-used mechanism involves the formation of a potential terminator structure upstream of the coding region, which functions in concert with activated cellular factors to either terminate at that site or allow transcription to proceed through it. In the latter case, the quiescent state allows transcription, and the binding of an activated factor activates the termination complex, causing transcription to abort. In the former case, the sequence of the nascent mRNA allows for the formation of a specific secondary structure capable of triggering the RNA polymerase to pause, prompting the release of both the polymerase and the RNA transcript from the DNA template [1]. In this case, the binding of a *trans*-acting factor alleviates this block, allowing transcriptional read through. This process is generally referred to as an attenuation or anti-termination mechanism of regulation. The distinction between the attenuation and anti-termination pathways is the end result of interactions between the terminator and the activated protein: is transcription terminated or allowed to continue? In the attenuation process, the regulatory protein, activated by the regulatory molecule, pauses the transcription at the terminator structure, which otherwise permits the read through of the transcription apparatus. The best example for this kind of regulation is the tryptophan biosynthetic operon

[2, 3]. In contrast to this mechanism, the anti-termination process requires the activated protein to bind to the pre-existing terminator structure, to allow the RNA polymerase to transcribe the downstream genes. An example of this is the BglG/SacY family of anti-termination proteins [4, 5] (Fig. 1). The target sequences for these anti-terminator proteins comprise either single or double stranded regions of their respective mRNAs.

In order to understand these mechanisms of attenuation and anti-termination in bacteria, there are only two protein-RNA complexes available for study. These are the TRAP attenuation protein and the LicT anti-termination protein; surprisingly, both are from *Bacillus subtilis*. TRAP regulates transcription of the *B. subtilis* tryptophan biosynthetic (*trp*) operon. TRAP is composed of 11 identical subunits arranged in a ring shaped molecule. Upon activation through the binding of up to 11 L-tryptophan molecules, TRAP binds to a RNA target consisting of 11 GAG and UAG repeats in the 5' leader region of the mRNA [6, 7]. TRAP binding inhibits the formation of an anti-terminator structure, and hence promotes the formation of a terminator hairpin, which halts transcription in the *trp* leader region (upstream of the structural genes). Although the binary and ternary complexes of TRAP revealed the important interactions with the L-tryptophan and RNA, respectively, the structure in the absence of ligand (apo-TRAP) is not available to reveal insights into any structural rearrangements that might occur upon ligand binding.

LicT is a member of the BglG-SacY family of anti-termination proteins, which regulate transcription of genes

*Address correspondence to this author at the RIKEN SPring-8 Center, Harima Institute, 1-1-1 Kouto, Sayo, Hyogo, 679-5148, Japan; Tel: +81 791 58 2838; Fax: +81 791 58 2826; E-mail: tsavel@spring8.or.jp

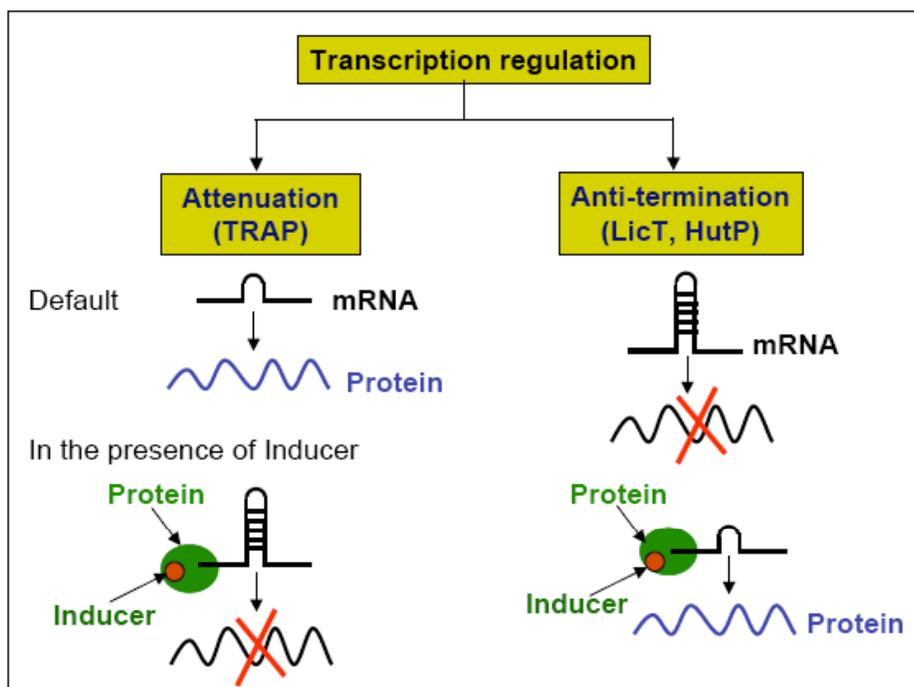


Fig. (1). Schematic representation of transcriptional regulation by an anti-termination or attenuation mechanism. The difference between the attenuation and anti-termination has been delineated. In the absence of ligand binding (non-activation), transcription proceeds in the case of attenuation mechanism; in contrast to this, anti-termination requires the activated protein for transcription.

involved in carbohydrate metabolism. These proteins bind to RNA targets known as RATs that consist of stem-loop structures, which overlap adjacent transcription terminators. Protein binding stabilizes the RAT structure, thereby inhibiting formation of the terminator. Apart from the N-terminal RNA binding domain (CAT), these proteins have two regulatory domains (PRD1 and PRD2), which contain sites for phosphorylation in response to carbohydrate availability. Recently, comparison of the native (inactive) and mutant (active) analyses of LicT showed massive tertiary and quaternary rearrangements of the entire regulatory domain. In the inactive state, a wide swing movement of PRD2 results in dimer opening and brings the phosphorylation sites to the protein surface. This movement is accompanied by additional structural rearrangements of both the PRD1-PRD1' interface and the CAT-PRD1 linker [8]. However, the structural information is available only for the N-terminal fragment with RAT and therefore the complete interactions of RAT with LicT were unobtainable [9]. Both of the available protein-RNA complexes lack the information necessary to understand the activation or the overall mechanism of anti-termination proteins. We chose HutP for our study to reveal the complete anti-termination mechanism; the detailed analysis of this mechanism is described in this review.

DESCRIPTION OF THE HutP

HutP, Histidine utilizing Protein, is one of the anti-terminator proteins in *Bacillus subtilis*. It is responsible for regulating the expression of the *hut* structural genes in response to changes in the intracellular levels of L-histidine [10, 11]. In the *hut* operon, HutP is located just downstream from the promoter, while the five other subsequent structural genes, *hutH*, *hutU*, *hutI*, *hutG* and *hutM*, are positioned far

downstream from the promoter [12-15]. In the presence of L-histidine and divalent metal ions, HutP binds to the nascent *hut* mRNA leader transcript. This allows the anti-terminator to form, thereby preventing the formation of the terminator and permitting transcriptional read through into the *hut* structural genes. In the absence of L-histidine and divalent metal ions or both, HutP does not bind to the *hut* mRNA, thus allowing the formation of a stem loop terminator structure within the nucleotide sequence located between *hutP* and the structural genes [11]. Similar to HutP, many regulatory proteins that involve allosteric regulation by small molecules to modulate their binding to the cognate mRNA, have been described for various operons [16-22]. These proteins must initially be activated by their specific ligands before they can function as anti-terminators/attenuators.

HutP is a 16.2 kDa protein consisting of 148 amino acid residues. HutP also exists in five other *Bacillus* species, including *B. anthracis* [23], *B. cereus* [24], *B. halodurans* [25], *B. thuringiensis* [26] and *Geobacillus kautophilus* [27] with 60% sequence identity. Sequence comparisons with the other *Bacillus* HutPs revealed that the C-terminal amino acid residues are more conserved than the N-terminal residues. Unlike other regulatory proteins, HutP not only requires its ligand (L-histidine), but also divalent metal ions for its activity.

ROLE OF L-HISTIDINE IN HutP

HutP requires L-histidine for binding to the *hut* mRNA [28]. Thus, L-histidine appears to allosterically control HutP-RAT interactions by modifying the conformational state of HutP. For the complete activation of HutP, a concentration of ~10 mM of L-histidine was sufficient for binding to the cognate RNA [28]. To obtain additional insights into the

requirement of L-histidine, we analyzed fifteen different L-histidine analogs. Using a filter binding assay, our goal was to determine not only the functional groups responsible for activation, but also to find the suitable analogs that exhibit higher affinity. Among the analogs tested, L-histidine β -naphthylamide (HBN) and L-histidine benzylester showed higher affinity (10-fold) over L-histidine. L-histidine methyl ester and L- β -imidazole lactic acid showed similar affinity as L-histidine (Kd \sim 300 nM), and urocanic acid, histamine, and L-histidinamide showed only weak activation. D-Histidine, imidazole-4-acetic acid, L-histidinol, α -methyl-DL-histidine, 1-methyl L-histidine, 3-methyl L-histidine, and 3-(thienyl)-L alanine failed to show any activation [28, 29].

Based on the analysis of the active analogs, we found that the imidazole group as well as the backbone moiety of L-histidine is essential for activation. Moreover, HutP activation by L-histidine is highly stereospecific since D-histidine prevented *hut* mRNA binding to HutP. This suggests that the correct positioning of the α -amino and carboxy moieties of L-histidine is essential for HutP activation.

ROLE OF DIVALENT METAL IONS FOR THE ACTIVATION OF HutP

Our previous analyses suggested that HutP binds to its cognate RNA only in the presence of L-histidine [28, 30]. To assess the ability of HutP to bind to mRNA, reactions were carried out in the presence of L-histidine (10 mM) and Mg²⁺ ions (10 mM). However, we do not know the importance of the metal ions in this anti-termination complex formation. To analyze the importance of the divalent metal ions in the formation of the HutP-L-histidine-RNA complex, we used a non-denaturing gel-shift assay to monitor the complex.

When MgCl₂ was omitted from the binding reactions, HutP failed to bind to *hut* mRNA. When 0.5 mM of MgCl₂ was incorporated in the binding buffer, we clearly observed anti-termination complex formation. The amount of complex formation increased further with higher metal ion and L-histidine concentrations, and displayed concentration dependence [31]. The metal ion Kd (489 μ M) value for the HutP-RNA interactions appeared to be more efficient (>10-fold) compared to the metal ions for other protein-RNA interactions, suggesting the existence of an efficient metal ion binding pocket. Therefore, HutP represents the first example of a single-stranded RNA-binding protein that requires metal ions for mediating RNA-protein interactions.

Several divalent metal ions (Mg²⁺, Ca²⁺, Mn²⁺, Zn²⁺, Co²⁺, Cd²⁺, Ba²⁺, Sr²⁺, Ni²⁺, Pb²⁺, Ag²⁺, Pt²⁺) were found to mediate HutP-RNA interactions and a few metal ions (Cu²⁺, Yd²⁺, Hg²⁺) that failed to support activity. Among the 12 divalent ions that support the interactions, Mn²⁺, Zn²⁺, and Cd²⁺ are found to be the most efficient, followed by Mg²⁺, Co²⁺, and Ni²⁺. Both the tested monovalent cations (Na⁺, K⁺) failed to mediate the HutP-RNA interactions even at 10 mM and 100 mM concentrations in the absence of divalent cations. These analyses conclude that divalent cations are mandatory for the interactions between HutP and its RNA and cannot be replaced with monovalent cations [32].

CHARACTERIZATION OF *hut* mRNA

hut mRNA appears to form two alternative secondary structures that depend upon the presence of activated HutP [30]. The regions that fold into alternative structures are located in between +489 to +572 of *hut* mRNA (Fig. 2). *hut* mRNA appears to form a stable terminator structure (+498 to

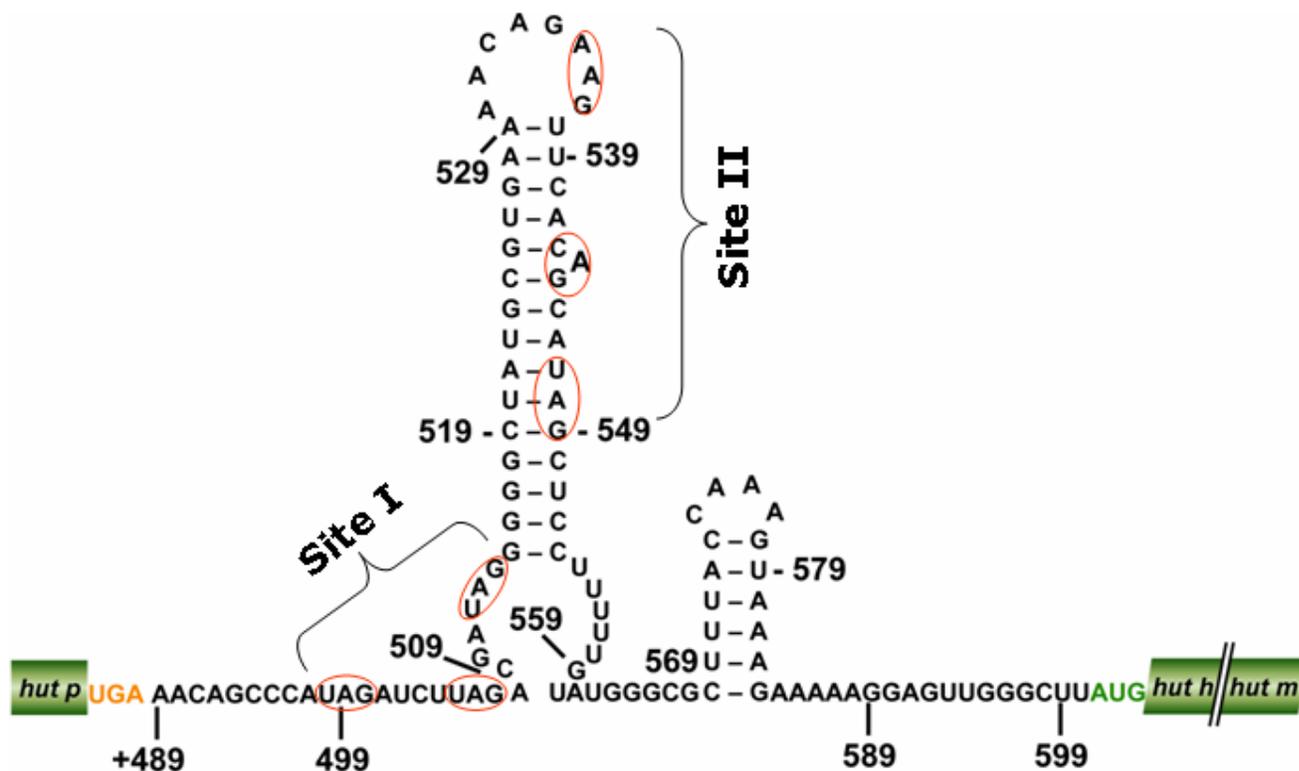


Fig. (2). The proposed *hut* mRNA terminator structure. The stop codon of the *hutP* gene and the start codon of the *hutH* gene are shown in orange and green, respectively. Within the HutP-recognition motif, the XAG (X indicates any base) motifs are highlighted with red box.

572) in the absence of activated HutP, whereas the region between +489 to +537 seems to form a destabilized RNA structure (ribonucleic anti-terminator, RAT) upon binding to the activated HutP [30]. Our biochemical [30] and recent structural studies [28] revealed that HutP forms a hexamer, as a trimer of dimers. Based on these results, a search was made for repeated sequences within the minimal 79-mer RAT element, and found three UAG repeats that are each separated by a 4-nucleotide spacer region. To determine whether this is the HutP binding region, we synthesized a short RNA (21mer-RAT) representing the region from +496 to +516 of the *hut* mRNA, and analyzed its binding to the activated HutP [28]. The equilibrium binding studies revealed that the 21-mer-RAT (5'-CAUAGAUCUUAGACGAUA GGG-3') binds to activated HutP as efficiently as the 79-mer RAT. Thus, the 21-mer RNA represents the minimal RNA element sufficient for HutP recognition. The bases located in the spacer regions in between the UAG motifs could be replaced with other bases, but the four nucleotide spacing was optimal for HutP recognition.

To ascertain the importance of each base as well as the tolerance for base substitutions in the UAG motif, we adopted an *in vitro* selection strategy along with site-specific mutational analyses [33]. Based on these experiments, we identified 5'-UUUAGNNNUAGNNNUAGUU-3' as the recognition motif, where N indicates any base. These analyses further suggested that each HutP monomer recognizes one UAG motif, where the first base (U) can be substituted with other bases, and the second and third bases (A and G) are required for the interactions. The 6-amino group of the A base and the 2-amino group of the G base are the most important groups for the protein-RNA interactions. Additionally, the 2'-OH of A in the UAG motif was found to be very important for HutP recognition, based on the 2'-deoxyribose base substitutional analysis at the binding motif [32]. The stoichiometric analysis of HutP-RNA interactions through the gel mobility shift assays confirmed that the hexameric HutP has two potential sites for RNA (1:2 ratio of HutP:RNA).

Based on the structural analysis of HutP with 21-mer RNA (described below) and our biochemical analyses suggesting for two potential binding sites in HutP, we reexamined of the entire intervening region between the *hutP* gene stop codon and the initiation codon of the first structural gene, *hutH*, revealed interesting phenomenon that there are two potential binding sites (site I, +498 to +514; site II, +535 to +549) in the intervening sequence, and each binding site consists of three XAG motifs (X denotes any base). To evaluate whether HutP binds to RNA with two XAG-rich sites (I and II), we chemically synthesized a model RNA (55-mer; 5'-UUUAGUUUUUAG UUUUUAGUUUUUUUUUUUUUUUUUUAGUUUUUUAGUUUUUUAGUU-3'), in which repeat XAG regions are joined by a linker region of 17 U's, similar to the length of the linker region present between sites I and II in the *hut* terminator RNA (Fig. 2). In order to clarify RNA-protein interactions, we carried out a gel-shift analysis and found that hexameric HutP binds to model RNA in a cooperative manner, forming a 1:1 ratio of RNA to protein complex, as compared to the 21-mer RNA (1:2 ratio). Based on this analysis, we presumed that both the upstream and downstream UAG-rich motifs are important, not only for binding to the specific residues within the hex-

americ HutP but also for causing significant conformational changes in the linker region.

Next, we analyzed the HutP binding sites within the 55-mer RNA using an in-line mapping assay, which relies on the structure-dependent spontaneous cleavage of RNA in the presence of divalent metal ions [34]. Our in-line mapping revealed that effective cleavage was observed in the XAG regions at both sites within the 55-mer RNA, but only in the presence of activated HutP (HutP-L-histidine-Mg²⁺). These sites are analogous to the UpU sites before the AG residues of three repeats of a 7-mer RNA (5'-(UpUpUpApGpUpU)₃-3'), which we used previously in our structural analysis [31]. In addition to the above sites, the linker region U₁₇ between the two UAG-rich regions was also cleaved efficiently, in the presence of activated HutP. These results suggest that the two XAG-rich sites (I and II) in the model 55-mer RNA are recognized specifically by the hexameric HutP, probably on its two (top and bottom) surfaces, and that the binding causes conformational changes near the UAG sites and the 17-nucleotide linker region. Further mapping studies using two kinds of RNA, having substitutions of the residues in the UAG motifs, either at 5'-end (site I) or 3'-ends (site II) revealed that the cleavage patterns observed in these RNAs were significantly different from 55-mer RNA having two XAG-rich regions. The absence of cleavages in these two RNAs reflects the inability of HutP modulation, suggesting that until the two UAG motifs are fully bound by HutP, the linker region may not undergo a significant conformational change. Moreover, these studies suggest that hexameric HutP can access two sites within the 55-mer RNA. The above in-line mapping studies were then extended to the wild-type *hut* terminator RNA, and these analyses revealed that the two XAG-rich sites in the *hut* terminator RNA are recognized by hexameric HutP, and for this, a 17-nucleotide linker length is sufficient for the placement of the next potential binding site [34]. We have confirmed these results by analyzing different RNA variants, with substitutions of their complementary residues in their XAG motifs. Finally, we have also confirmed the importance of the two XAG-rich regions within the *hut* terminator RNA in the overall anti-termination process by *in vivo hut* anti-termination assays [34]. Interestingly, many of these *hut* terminator RNAs in *Bacillus* species (*B. Subtilis*, *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. halodurans*) have two conserved XAG-rich regions flanking the linker region of 12-17 nucleotides.

CRYSTAL STRUCTURES OF HutP

The *in vitro* and *in vivo* functional analysis of the HutP anti-termination complex suggest that L-histidine, divalent metal ions, and an RNA containing the three XAG repeating motifs separated by four spacer nucleotides are mandatory for complex formation. In order to reveal snap-shots of the complete mechanism of HutP, we solved the crystal structures of the uncomplexed HutP [31], HutP-Mg²⁺ [31], HutP-L-histidine β-naphthylamide (HBN, an analog of L-histidine) [28, 32], HutP-L-histidine-Mg²⁺ [31] and HutP-L-histidine-Mg²⁺-RNA (21mer & 55-mer) [31, 34-37]. The asymmetric unit (asu) of uncomplexed HutP (PDB id, 1WPS), HutP-Mg²⁺ (PDB id, 1WPT), binary (HutP-HBN) (PDB id, 1VEA), and quaternary complex (HutP-L-histidine-Mg²⁺-21mer RNA) (PDB ids, 1WMQ, 1WPU), contain two HutP

molecules; whereas the ternary complex (HutP-L-histidine- Mg^{2+}) (PDB id, 1WPV) and quaternary complex with 55-mer RNA (PDB id, 2GZT) contains 3 HutP molecules (Fig. 3). All of these asu dimers are related by a 3-fold crystallographic symmetry and the trimers are related by a 2-fold crystallographic symmetry, in order to form a hexameric structures. The hexamer formation of HutP is important for its function and is consistent with the gel filtration assays. Each monomeric HutP molecule consists of four α -helices and four β -strands, arranged in the order α - α - β - α - α - β - β in the primary structure, and the four antiparallel β -strands form a β -sheet in the order β 1- β 2- β 3- β 4, with two α -helices each on the front and on the back. Although HutP belongs to the α/β family of proteins, a thorough search of the PDB [38] and SCOP [39] databases revealed that the HutP fold is novel and not related to any of the existing folds [28]. This result was expected based on the sequence analysis of HutP with RAT sequences.

Apo-Form of HutP

Superposition of uncomplexed HutP and HutP- Mg^{2+} reveals that both exhibit almost similar conformations. The RMS deviation of the superposition between these two structures was 0.80 Å. At the dimer interface, an open hydrophobic pocket is formed and stabilized by a salt bridge between the monomers of HutP, Glu81 of Molecule A and Arg88 of Molecule B. However, Mg^{2+} ions are not bound in the HutP- Mg^{2+} complex [31, 32] (Fig. 3a).

Binary Complex of HutP

To determine the crystal of structure of HutP with L-histidine, an analog of L-histidine, L-histidine β -naphthylamide (HBN), was used since it showed the highest affinity for activation. HutP was co-crystallized with HBN and the structure was solved [28, 35]. The overall structure of this binary complex resembles that of the uncomplexed HutP (RMSD, 0.77) and HutP- Mg^{2+} (RMSD, 0.60) structures (Fig. 3b). HBN bound between the monomers of HutP; the imidazole group of L-histidine was buried in a hydrophobic pocket created by the two adjacent protein molecules. The imidazole ring amide nitrogen ND1 hydrogen bonded with the side chain of Glu81 of Molecule B and the same Glu81 bonded to Arg88 through a salt bridge. The imidazole ring nitrogen (NE2) was hydrogen bonded to the backbone moiety of Glu139 of Molecule A; it also made a hydrophobic contact with the phenyl ring of Phe141 of Molecule A. Mutational analyses of these critical residues suggested that the Glu81 and Arg88 were important for the analog interactions.

Ternary Complex of HutP

The asymmetric unit of the ternary complex (HutP-L-histidine- Mg^{2+}) is shown in Fig. (3c). The dimer of the ternary complex contained 2 L-histidines and 2 Mg^{2+} ions, located on either side of the dimer interface [31, 32]. The L-histidine imidazole ring projects downwards into the solvent, with the imidazole nitrogen (NE2) hydrogen bonding to a water molecule and the amide nitrogen (ND1) hydrogen bonding to the sidechain oxygen of Tyr69 of the neighboring dimer within the HutP hexamer. In contrast to the HutP-HBN complex, the amino (N) and carboxyl (CO and OXT) groups of the L-histidine backbone in the ternary complex

have many interactions. For instance, the carboxyl group of the L-histidine backbone in the ternary complex forms a typical salt bridge with the guanidyl group of Arg88. In the

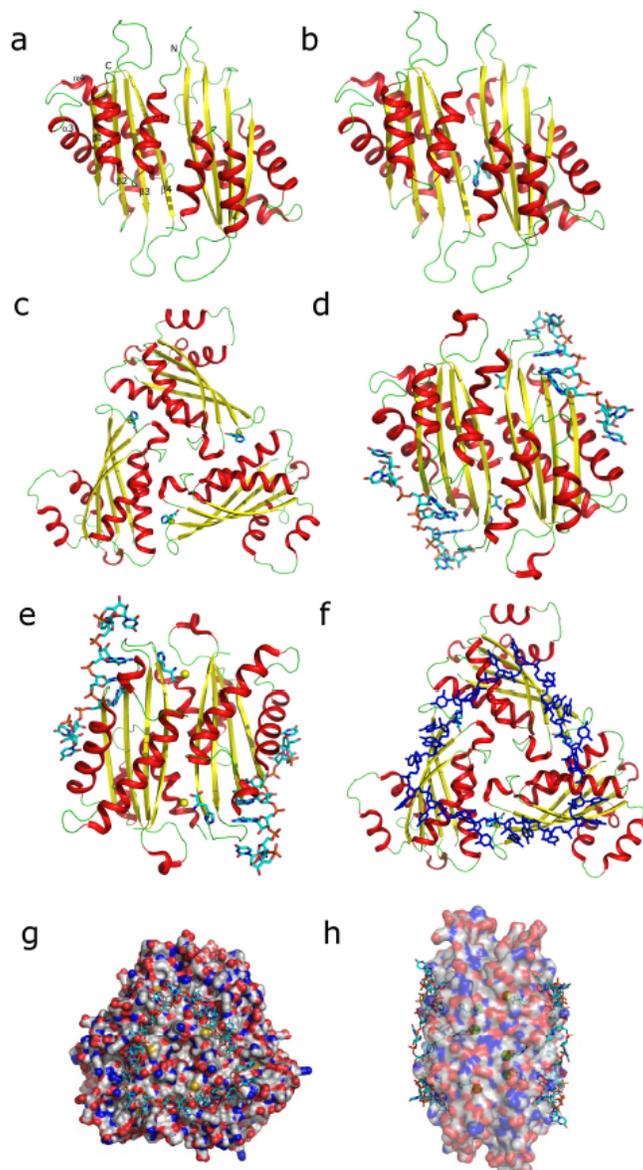


Fig. (3). Crystal structures of HutP. (a) Structure of the uncomplexed HutP or HutP- Mg^{2+} . (b) Binary complex of HutP with L-histidine complex. (c) Ternary complex of HutP (HutP-L-histidine- Mg^{2+}). (d,e) Front and back view of quaternary complex of HutP (HutP-L-histidine- Mg^{2+} -21mer RNA). (f) Quarternary complex of HutP with 55-mer RNA. (g,h) The biological assembly of HutP is hexamer shown in surface representation. The bound Mg^{2+} ions in ternary and quaternary complexes are represented by a cpk model. The bound RNA and the analog (HBN in binary and L-histidine in ternary and quaternary) complexes shown in a ball-and-stick model. The asymmetric unit of HutP molecules represented in the figures a-f. The figure panels were prepared using Pymol [42].

HutP-HBN complex, this guanidyl group forms a typical salt bridge with the carboxyl group of Glu81 of Molecule A, which is important for the formation of the hydrophobic

pocket [28]. The amino group as well as the carboxyl group of L-histidine coordinates with the Mg^{2+} ion, in the ternary complex. The Mg^{2+} ion is involved in the formation of a typical six-coordination sphere. Of the six coordinations, three were with the imidazole nitrogens of His138, His73, and His77. The latter two histidines were those from the neighboring dimer. The sixth coordination of the Mg^{2+} ion was that with a water molecule, which was anchored by a hydrogen bond with the side chain of Glu81 of Molecule A. To further substantiate the metal ion binding site residues on HutP, the residues that interact with the metal ions, His73, His77, and His138, were individually substituted with Ala, and the protein-RNA interactions were evaluated by a gel mobility shift assay [31]. These assays clearly showed that all three residues in the histidine cluster were important for coordination of the Mg^{2+} ion. Thus, the Mg^{2+} ion helps to mediate the interaction between HutP and L-histidine, along with the residues within the specific binding site in the dimer and the dimer-dimer interface of HutP. Recent crystal structures complexes, HutP-L-histidine- Mn^{2+} and HutP-L-histidine- Ba^{2+} , reveal the fact that the Mg^{2+} binding site can also accommodate different divalent cations, with ionic radii ranging from 0.72-1.35 Å, as suggested by the biochemical analysis [32].

Quaternary Complex of HutP with 21-mer RNA

The asymmetric unit of quaternary complex contains one homodimer of HutP, two L-histidines, two Mg^{2+} ions, and two 7-nucleotide fragments of the 21-mer RNA, which are related by a non-crystallographic two fold axis [31, 35] (Fig. 3d and e). In the HutP dimer, L-histidine and Mg^{2+} are associated with each other, and are located in the dimer interface as observed in the HutP-L-histidine- Mg^{2+} complex. In both quaternary complexes, the bound L-histidine and Mg^{2+} ions are recognized on one side of HutP (Fig. 3e) and the RNA is recognized on the other side (Fig. 3d). However, the Mg^{2+} ion does not bind directly to the RNA. This observation is in contrast with the role of metal ions reported in the past; for example, in the L11-RNA protein complexes, the metal ions occupy crucial locations to stabilize the sharp turns of the junctions, strengthening the overall structure of the four-way junction [40]. The bound 21-mer RNA are recognized on both the top and bottom surfaces of the HutP hexamer in a novel triangular fashion (Fig. 3g and h). Thus, the HutP hexamer recognizes two 21-mer RNA molecules, which is consistent with our previous biochemical studies as suggested it predominantly forms 1:2 (Protein:RNA) complex ratios [28, 33].

Each HutP monomer recognizes one unit of the 7-nucleotide fragments from the 21-mer RNA (5'-UUUAGUU-3')₃. In the 7-nucleotide RNA, the central A4-G5 forms an extensive hydrogen bonding network with HutP. The U7 interacts with the neighboring dimer within the HutP hexamer. Bases U1, U2, U3, and U6 were partially disordered and showed no interactions with HutP. This analysis suggested that the U3 base is replaceable with other RNA bases and is not important for proper function, which is consistent with our biochemical analysis [32]. The overall RNA structure adopts an extended A-RNA conformation, with C3' endo sugar conformations for all of the residues. The central bases A4 and G5 are stacked upon each other, and the remaining bases probably serve as spacers without

much interaction with the protein molecule; their basic requirement is apparently to place the next UAG binding site on the hexameric HutP.

No direct protein-RNA phosphate interactions were observed in the HutP ternary complex. The 2'-OH groups of the A4 and U7 sugars hydrogen bonded with the side chains of Thr99 and Thr56, respectively. These two interactions involving the 2'-OH are critical for the protein-RNA interactions, and is consistent with the deoxyribo nucleotide substitutional analysis [32] as well as the mutational analysis of HutP. This explains the ability of HutP to distinguish between RNA and DNA. All of the hydrogen bonding interactions in this protein-RNA complex are mainly directed toward the base and the sugar, and not the phosphate backbone; this readily explains the sequence-dependent nature of the recognition. To further substantiate the RNA binding site residues on HutP, the critical residues that interact with the RNA—Glu55, Thr56, Thr99, Thr128 and Glu137—were substituted with Ala, and the protein-RNA interactions were evaluated using both the gel mobility shift and filter binding assays. The Thr99Ala and Glu137Ala mutant proteins failed to bind to the RNA, the Thr128Ala mutant was reduced by 15-fold, and the Glu55Ala and Thr56Ala mutants showed comparable affinities to the wild-type HutP (Kd 95 ± 14 nM). These studies clearly showed that residues Thr99 and Glu137 were highly responsible for RNA binding [31]. These analyses further suggested that the 2'-OH of the A4 ribose interaction with Thr99 was more important than the interaction of the 6-amino group of the A4 base with Thr128. This was also the reason why A4 could be substituted with U with 30- fold less affinity [33], owing to the loss of the 6-amino interactions with Thr128. These three residues (Thr99, Thr128, and Glu137) are completely responsible for the RNA base-specific recognition by HutP. The other two residues, Glu55 and Thr56, have additional contacts with the spacer bases, but do not affect the overall binding or the complex structure formation. This analysis also supports our previous biochemical results.

Quaternary Complex of HutP with 55-mer RNA

Recently, we solved the crystal structure of the complex (HutP-L-his- Mg^{2+} -55-mer RNA; PDB id, 2GZT) [34]. This 55-mer complex belongs to the monoclinic space group C2, with three molecules in the asymmetric unit (Fig. 3f). The overall hexameric structure of HutP and the binding mode with L-histidine, Mg^{2+} and RNA are essentially the same as those described for the previous quaternary complex with 21-mer RNA, with an r.m.s.d. of 0.17Å between the Ca pairs. However, it should be noted that the hexamer is created by different crystallographic symmetries. In the 55-mer complex, the three monomers within the trimer (asu) are related by non-crystallographic three-fold symmetry, and the two trimers are related by crystallographic two-fold symmetry. In contrast, in the 21-mer RNA complex, the three monomers are related by crystallographic three-fold symmetry, and the two trimers are related by non-crystallographic two-fold symmetry. The bound RNA is also related by two-fold and three-fold symmetries. The non-crystallographic three-fold symmetry in the 55-mer RNA complex allowed us to determine the entire structure of the 21-mer nucleotide fragment of the 55-mer RNA, thus providing more informa-

tion than the 21-mer RNA complex, in which only a 7-mer was determined, and the bases of the spacer nucleotides were somewhat disordered. The r.m.s.d. among U1-U7, U8-U14 and U15-U21 in the 55-mer RNA complex is within 0.44Å. A continuous electron density along the fragment U1-U21 is clearly visible, but that of the linker fragment U22-U34 is absent, except for the phosphate of U22. The electron densities of the 21-mer fragments of U1-U21 and U35-U55 are visible, because they are superposed well, but in the linker region those of U22 and the following nucleotides, and those of U34 and the preceding nucleotides are not visible, suggesting they are not superposed. Therefore, the electron densities of the linker region disappear by symmetry-averaging. Alternatively, the absence of electron density in this region is probably due to the large disorder of the linker, reflecting the high mobility of the linker region. This is consistent with the in-line mapping observation, in which the linker region undergoes a significant conformational change in the presence of activated HutP. The full-length of the RNA in the crystals were also confirmed by dissolving the quaternary complex (HutP-L-his-Mg²⁺-55-mer RNA) crystals in water, after washing them a few times in reservoir solution, and then isolated and analyzed the RNA, using reverse transcriptase. These analyses indicated that the RNA present in the crystals was, indeed, the 55-mer RNA [34].

In the 55-mer RNA complex, the bases of U1, U2, U6, U8, U9, U13, U15, U16 and U20 are exposed to the solvent. Although the bases of U3, U10 and U17 have no direct hydrogen bonds to HutP, the O2 atoms of these bases closely contact the N7 atoms of the adjacent bases A4, A11 and A18, respectively. Thus, the bases of the spacer U nucleotides weakly contact HutP, suggesting that these interactions may be non-specific, consistent with the finding that the spacer nucleotides are sequence-independent. Nevertheless, the quaternary complex of HutP with 55-mer RNA revealed, consistent with the in-line mapping and gel-shift analyses, which HutP indeed binds to two sites within the 55-mer model RNA, using its dual-RNA-binding surfaces. Thus, the binding manner of the 55-mer model RNA to HutP is expected to reflect the actual binding of the native *hut* terminator on the activated HutP.

ALLOSTERIC ACTIVATION OF HutP

The overall structure of the HutP monomer has almost similar secondary structure in all of our solved HutP structures. However, the HutP-L-histidine-Mg²⁺ complex showed significant conformational changes when it was superimposed on the other structures: uncomplexed HutP (RMSD of 3.1 Å), HutP-Mg²⁺ (RMSD of 2.73 Å), and HutP-HBN (RMSD of 3.15 Å). The conformational changes observed

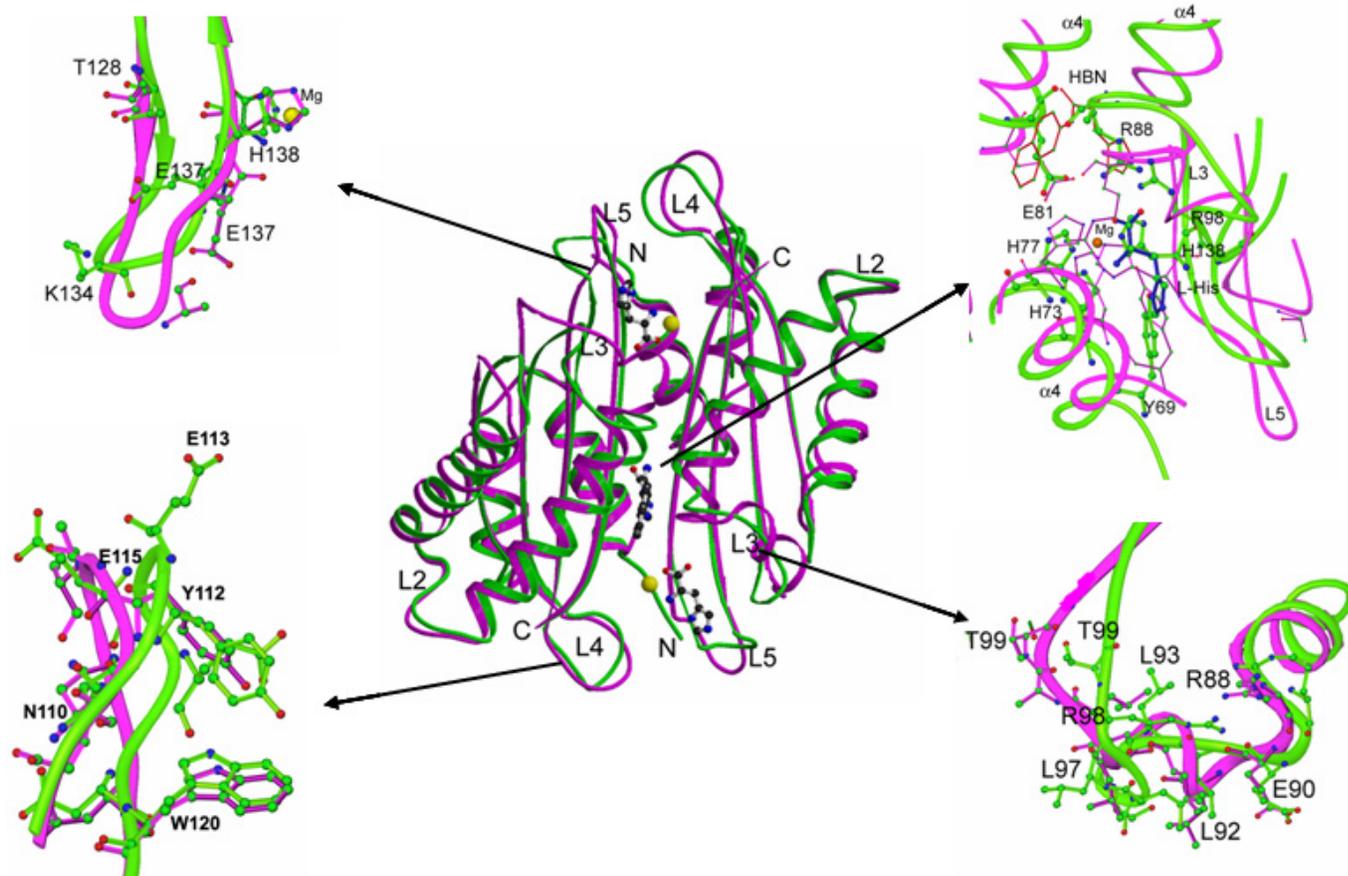


Fig. (4). Allosteric activation of HutP. Superposition of two crystal structures (binary complex in pink; quaternary complex in green), showing the conformational changes at the L-histidine binding site and the loop regions L3, L4, and L5 regions. The L-histidine β -naphthylamide (HBN), and L-histidine are shown in a ball-and-stick model, and the Mg²⁺ ions are represented by a cpk model. Significant conformational changes observed in the L-histidine binding site, loop L3, L4, and L4 are zoomed for ease viewing. The figure was prepared using Ribbons [43].

were especially at the L-histidine binding site and the loops, L3, L4 and L5 (Fig. 4). The uncomplexed HutP and HutP-Mg²⁺ structures clearly showed the existence of an open hydrophobic pocket at the dimer interface. The HutP-HBN complex revealed that the imidazole group of the bound L-histidine ligand was completely buried inside the hydrophobic pocket and established a hydrogen bonding network with the surrounding protein residues. In contrast to the previous structures, the HutP-L-histidine-Mg²⁺ complex distinctly shows the existence of two Mg²⁺-L-histidine interactions on either side of the dimer interface. Based on this, it appears that once the Mg²⁺ ion is bound by the HutP-L-histidine complex, the Mg²⁺ ion may pull the L-histidine out of the hydrophobic pocket, binding to it tightly. In this manner, the Mg²⁺-bound L-histidine could move along the dimer interface and reside approximately 12 Å away from the pocket. This movement of L-histidine is apparently linked to the movement of Arg88 in the opposite direction, leading to the disruption of the hydrophobic pocket formed by the salt bridge between Arg88 and Glu81 in the HutP-HBN complex, and the formation of a new salt bridge between Arg88 and Arg98. This drastic rearrangement of Arg98 is accompanied by a large change in the C_α position of the next residue, Thr99, of loop L3. On the other end, the reorientation of the His138 imidazole ring to coordinate with the Mg²⁺ ion causes a large conformational change in the local backbone chain, particularly in the torsion angles around His138 and the next residue Glu137. Also the backbone oxygen of Glu137 is bonded to the hydroxyl group of Try112. This changes the orientation of the sidechain of Glu137. These changes in HutP might have been initiated by the Mg²⁺ ion and transmitted along the backbone chain to Thr128 through

the L5 loop. In this ternary complex, the metal ion and L-histidine mutually interact and facilitate the required overall structural rearrangement, because even if one of the two components (metal ion or L-histidine) is absent, HutP fails to bind to the RNA to form an anti-termination complex [31]. The recognition of L-histidine must be the first step in this process, because until the imidazole ring of the L-histidine ligand is verified to be in the hydrophobic pocket, the HutP is not ready for the Mg²⁺ ion-mediated structural rearrangement. In the quaternary complexes, the protein did not exhibit further conformational changes upon binding to the RNA. Thus, L-histidine and Mg²⁺ ions are together responsible for the allosteric activation of HutP in the HutP-L-histidine-Mg²⁺ complex, especially the conformational changes at the L-histidine binding site and Loops L3 and L5. The loop L3 is responsible for critical interactions with 2'-OH of base A4 and the loop L5 is responsible for the interactions with the 6-amino of base A4 and the 2-amino of base G5 in quaternary complex [31, 34].

REVELATION OF ANTI-TERMINATION MECHANISM MEDIATED BY HutP

Based on the analyses of the various biochemical (*in vitro* and *in vivo*) and structural studies of HutP, a complete model is revealed for the anti-termination of HutP [31, 41] (Fig. 5). HutP forms a hexamer in the presence or absence of ligands. Thus, HutP exists as a hexamer that forms a hydrophobic pocket in the center of each HutP dimer (step A). In the next step, the incoming residue is verified in the pocket, recognizes only the imidazole-containing residue (L-histidine) from the others present, and then interacts exten-

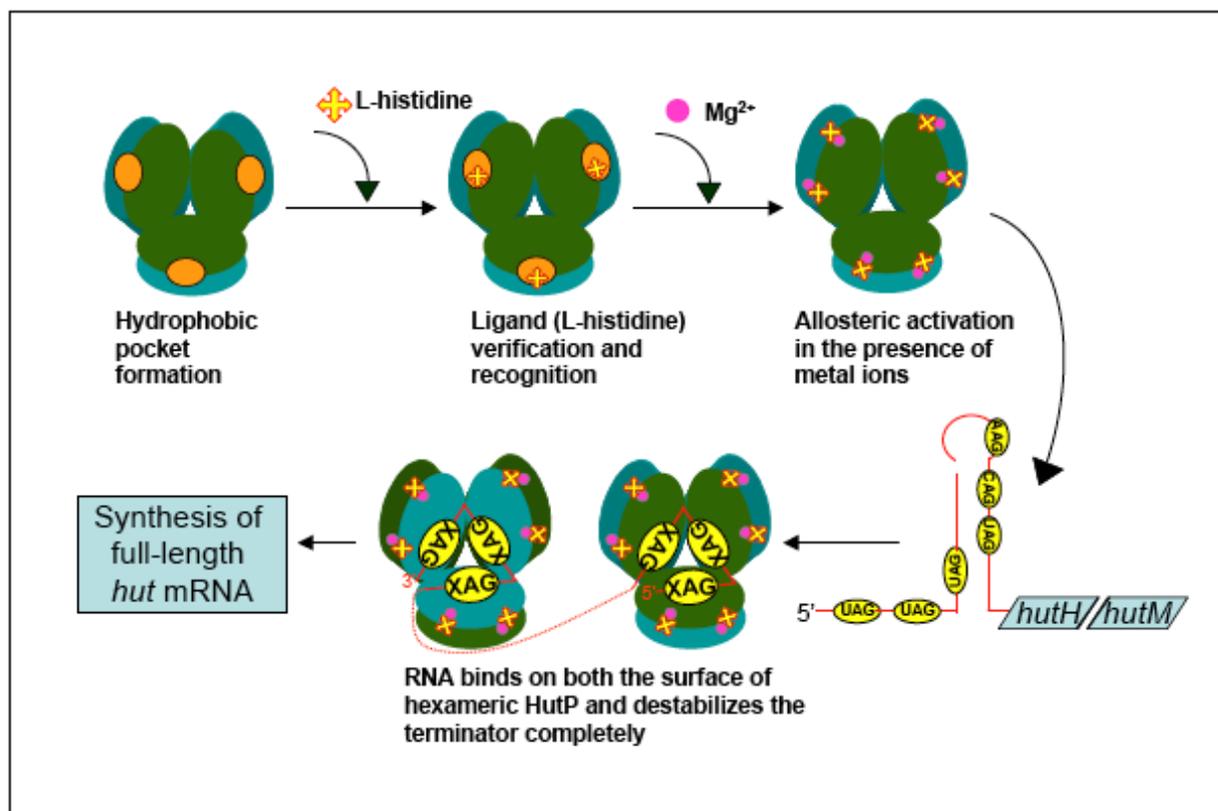


Fig. (5). Schematic representation of the HutP mediated anti-termination mechanism.

sively with the HutP residues in the pocket (step B). Once the L-histidine interactions have been completed in the pocket, HutP undergoes allosteric activation in the presence of Mg^{2+} ions or other divalent metal ions, producing structural rearrangements at the L-histidine binding site and loops L3 and L5 (step C). This causes L3 and L5 specifically to move and reorient the residues Arg88, Arg98, Thr99, Thr126, Glu137 and H138 for RNA recognition. Once HutP undergoes these conformational changes caused by the presence of L-histidine and Mg^{2+} ions, it is then able to recognize the specific chemical groups of the bases within the *hut* mRNA, without undergoing any further structural rearrangement of the protein (step D); then the bound RNA might undergo a conformational change in a novel triangular fashion. Among the two XAG-rich sites found the RNA, the I XAG-rich site binds to one surface of hexameric HutP, and the following linker nucleotides carry the II XAG-rich site to the other surface of HutP, so as to bind in the same fashion as in site I. As a result, the *hut* terminator appears to destabilize completely and wrap the dual RNA-binding surfaces of the HutP hexamer in a 5' to 3' direction.

CONCLUSIONS

Many RNA-binding proteins are known to play a major role in the regulation of gene expression through attenuation or anti-termination mechanisms by targeting the nascent RNA only in the presence of their cognate ligands. However, no structural insights were available to understand the anti-termination mechanism in each step. In order to rectify this, the present review summarizes the recent structural and biochemical analysis of HutP and revealed how HutP undergoes a conformational change in each step (apo-HutP, HutP- Mg^{2+} , HutP-L-histidine, HutP- Mg^{2+} -L-histidine, HutP- Mg^{2+} -L-histidine-RNA (21-mer and 55-mer RNA)) to form an active anti-termination complex. In conclusion, HutP represents the best model to understand the anti-termination mechanism, based on the existing structural and biochemical results.

ACKNOWLEDGEMENTS

The author thanks Dr. P.K.R. Kumar and Dr H. Mizuno for their kind encouragement and support.

REFERENCES

- Richardson JP, Greenblatt J. In *Escherichia coli* and *salmonella*: Cellular and Molecular Biology. In: Neidhardt FC, Curtiss III R, Ingraham JL, Eds. American society for Microbiology, Washington, D. C. 1996; pp. 822-48.
- Yanofsky C. Transcription attenuation: once viewed as a novel regulatory strategy. *J Bacteriol* 2000; 182: 1-8.
- Gollnick P, Babitzke P. Transcription attenuation. *Biochim Biophys Acta* 2002; 1577: 240-50.
- Henkin TM. Control of transcription termination in prokaryotes. *Annu Rev Genet* 1996; 30: 35-57.
- Rutberg B. Antitermination of transcription of catabolic operons. *Mol Microbiol* 1997; 23: 413-21.
- Antson AA, Dodson EJ, Dodson G, Greaves RB, Chen XP, Gollnick P. Structure of the *trp* RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* 1999; 401: 235-42.
- Antson AA, Otridge J, Brzozowski AM, *et al.* The structure of *trp* RNA-binding attenuation protein. *Nature* 1995; 374: 693-700.
- Graille M, Zhou CZ, Brechot VR, Collinet B, Declerck N, Van Tilbeurgh H. Activation of the LicT transcriptional antiterminator involves a domain swing/lock mechanism provoking massive structural changes. *J Biol Chem* 2005; 280: 14780-89.
- Yang Y, Declerck N, Manivel X, Aymerich S, Kochoyan M. Solution structure of the LicT-RNA antitermination complex: CAT clamping RAT. *EMBO J* 2002; 21: 1987-97.
- Oda M, Katagai T, Tomura D, Shoun H, Hoshino T, Furukawa K. Analysis of the transcriptional activity of the *hut* promoter in *Bacillus subtilis* and identification of a *cis*-acting regulatory region associated with catabolite repression downstream from the site of transcription. *Mol Microbiol* 1992; 6: 2573-82.
- Wray LV Jr, Fisher SH. Analysis of *Bacillus subtilis* *hut* operon expression indicates that histidine-dependent induction is mediated primarily by transcriptional antitermination and that amino acid repression is mediated by two mechanisms: regulation of transcription initiation and inhibition of histidine transport. *J Bacteriol* 1994; 176: 5466-73.
- Chasin LA, Magasanik B. Induction and repression of the histidine-degrading enzymes of *Bacillus subtilis*. *J Biol Chem* 1968; 243: 5165-78.
- Kimhi Y, Magasanik B. Genetic basis of histidine degradation in *Bacillus subtilis*. *J Biol Chem* 1970; 245: 3545-8.
- Oda M, Sugishita A, Furukawa K. Cloning and nucleotide sequences of histidase and regulatory genes in the *Bacillus subtilis* *hut* operon and positive regulation of the operon. *J Bacteriol* 1988; 170: 3199-205.
- Yoshida K, Sano H, Seki S, Oda M, Fujimura M, Fujita Y. Cloning and sequencing of a 29 kb region of the *Bacillus subtilis* genome containing the *hut* and *wapA* loci. *Microbiology* 1995; 141: 337-43.
- Houman F, Diaz-Torres MR, Wright A. Transcriptional antitermination in the *bgl* operon of *E. coli* is modulated by a specific RNA binding protein. *Cell* 1990; 62: 1153-63.
- Aymerich S, Stenmetz M. Specificity determinants and structural features in the RNA target of the bacterial antiterminator proteins of the *BglG/SacY* family. *Proc Natl Acad Sci USA* 1992; 89: 10410-4.
- Babitzke P, Yanofsky C. Reconstitution of *Bacillus subtilis* *trp* attenuation *in vitro* with TRAP, the *trp* RNA-binding attenuation protein. *Proc Natl Acad Sci USA* 1993; 90: 133-37.
- Arnaud MD, Debarbouille M, Rapoport G, Saier MH Jr, Reizer J. *In vitro* reconstitution of transcriptional antitermination by the SacT and SacY proteins of *Bacillus subtilis*. *J Biol Chem* 1996; 271: 18966-72.
- Lu Y, Turner RJ, Switzer RL. Function of RNA secondary structures in transcriptional attenuation of the *Bacillus subtilis* *pyr* operon. *Proc Natl Acad Sci USA* 1996; 93: 14462-67.
- Alpert CA, Siebers U. The *lac* operon of *Lactobacillus casei* contains lacT, a gene coding for a protein of the *BglG* family of transcriptional antiterminators. *J Bacteriol* 1997; 179: 1555-62.
- Glatz E, Nilsson RP, Rutberg L, Rutberg B. A dual role for the *Bacillus subtilis* *glpD* leader and the GlpP protein in the regulated expression of *glpD*: antitermination and control of mRNA stability. *Mol Microbiol* 1996; 19: 319-28.
- Read TD, Peterson SN, Tourasse N, *et al.* The genome sequence of *Bacillus anthracis* Ames and comparison to closely related bacteria. *Nature* 2003; 423: 81-6.
- Ivanova, Sorokin NA, Anderson I, *et al.* Genome sequence of *Bacillus cereus* and comparative analysis with *Bacillus anthracis*. *Nature* 2003; 423: 87-91.
- Takami H, Nakasone K, Takaki Y, *et al.* Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res* 2000; 28: 4317-31.
- Brettin TS, Bruce D, Challacombe JF, *et al.* Complete genome sequence of *Bacillus thuringiensis* 97-27. Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
- Takami H, Takaki Y, Chee GJ, *et al.* Thermoadaptation trait revealed by the genome sequence of thermophilic *Geobacillus kaustophilus*. *Nucleic Acids Res* 2004; 32: 6292-303.
- Kumarevel TS, Fujimoto Z, Karthe P, Oda M, Mizuno H, Kumar PKR. Crystal structure of activated HutP; an RNA binding protein that regulates transcription of the *hut* operon in *Bacillus subtilis*. *Structure* 2004; 12: 1269-80.
- Kumarevel TS, Mizuno H, Kumar PKR. Allosteric activation of HutP protein, that regulates transcription of *hut* operon in *Bacillus subtilis*, mediated by various analogs of histidine. *Nucleic Acids Res Suppl* 2003; 3: 199-200.
- Oda M, Kobayashi N, Ito A, Kurusu Y, Taira K. *cis*-acting regulatory sequences for antitermination in the transcript of the *Bacillus subtilis* *hut* operon and histidine-dependent binding of HutP to the

- transcript containing the regulatory sequences. *Mol Microbiol* 2000; 35: 1244-54.
- [31] Kumarevel TS, Mizuno H, Kumar PKR. Structural basis of HutP-mediated anti-termination and roles of the Mg²⁺ ion and L-histidine ligand. *Nature* 2005; 434: 183-91.
- [32] Kumarevel TS, Mizuno H, Kumar PKR. Characterization of the metal ion binding site in the anti-terminator protein, HutP, of *Bacillus subtilis*. *Nucleic Acids Res* 2005; 33: 5494-502.
- [33] Kumarevel TS, Gopinath SCB, Nishikawa S, Mizuno H, Kumar PKR. Identification of important chemical groups of the *hut* mRNA for HutP interactions that regulate the *hut* operon in *Bacillus subtilis*. *Nucleic Acids Res* 2004; 32: 3904-12.
- [34] Gopinath SCB, Balasubramanian D, Kumarevel TS, Misono TS, Mizuno H, Kumar PKR. Insights into anti-termination regulation of the *hut* operon in *Bacillus subtilis*: importance of the dual RNA-binding surfaces of HutP. *Nucleic Acids Res* 2008; 36: 3463-73.
- [35] Kumarevel TS, Fujimoto Z, Padmanabhan B, *et al.* Crystallization and preliminary X-ray diffraction studies of HutP protein: an RNA-binding protein that regulates the transcription of *hut* operon in *Bacillus subtilis*. *J Struct Biol* 2002; 138: 237-40.
- [36] Kumarevel TS, Fujimoto Z, Mizuno H, Kumar PKR. Crystallization and preliminary X-ray diffraction studies of the metal-ion-mediated ternary complex of the HutP protein with L-histidine and its cognate RNA. *BBA- Proteins Proteomics* 2004; 1702: 125-28.
- [37] Kumarevel TS. Structural insights of HutP-mediated regulation of transcription of the *hut* operon in *Bacillus subtilis*. *Biophys Chem* 2007; 128: 1-12.
- [38] Berman HM, Battistuz T, Bhat TN, *et al.* The Protein Data Bank. *Acta Crystallogr D* 2002; 58: 899-907.
- [39] Murzin AG, Brenner SE, Hubbard T, Chothia C. SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 1995; 247: 536-40.
- [40] Wimberly BT, Guymon R, McCutcheon JP, White SW, Ramakrishnan V. A detailed view of a ribosomal active site: the structure of the L11-RNA complex. *Cell* 1999; 97: 491-502.
- [41] Kumar PKR, Kumarevel TS, Mizuno H. Structural basis of HutP-mediated transcription anti-termination. *Curr Opin Struct Biol* 2006; 16: 18-26.
- [42] Delano WL. The pymol molecular graphics system. Delano Scientific, San Carlos, CA: USA 2002.
- [43] Carson M. Ribbons. *Methods Enzymol* 1997; 277: 493-505.

Received: January 13, 2009

Revised: February 19, 2009

Accepted: February 19, 2009

© Thirumananseri Kumarevel; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.